DOI:
10.1214/13-BA849

Inference in Two-Piece Location-Scale Models with Jeffreys Priors

F. J. Rubio and M. F. J. Steel*

University of Warwick, Department of Statistics, Coventry, CV4 7AL, UK

Abstract

This paper addresses the use of Jeffreys priors in the context of univariate three-parameter location-scale models, where skewness is introduced by differing scale parameters either side of the location. We focus on various commonly used parameterizations for these models. Jeffreys priors are shown to lead to improper posteriors in the wide and practically relevant class of distributions obtained by skewing scale mixtures of normals. Easily checked conditions under which independence Jeffreys priors can be used for valid inference are derived. We also investigate two alternative priors, one of which is shown to lead to valid Bayesian inference for all practically interesting parameterizations of these models and is our recommendation to practitioners. We illustrate some of these models using real data.

Key words and phrases: Bayesian inference; noninformative prior; posterior existence; scale mixtures of normals; skewness

*Corresponding author: Mark Steel, Department of Statistics, University of Warwick, Coventry, CV4 7AL, U.K. (Email: M.F.Steel@stats.warwick.ac.uk). We are grateful to two referees and the Associate Editor for insightful comments. Javier Rubio acknowledges research support from Conacyt (Mexico).

1 Introduction

The use of skewed distributions is an attractive option for modeling data presenting departures from symmetry. Several mechanisms to obtain skewed distributions by appropriately modifying symmetric distributions have been presented in the literature (Azzalini, 1985; Fernández and Steel, 1998; Mudholkar and Hutson, 2000).

We focus on the simple univariate location-scale model where we induce skewness by the use of different scales on both sides of the mode and only distinguish three scalar parameters. We investigate Bayesian inference using Jeffreys priors in this simple setting. Despite the simplicity of these models they often fit observed data quite well, and have been used recently in a wide variety of applied contexts, such as genetics, biology, hydrology, economics, finance, medicine, agriculture and marketing (Purdom and Holmes, 2005; Trindade et al., 2010; Rubio and Steel, 2011; Punathumparambath et al., 2012). For example, they are used for the widely discussed probability forecasts of gross domestic product and inflation produced by the Bank of England and the Sveriges Riksbank (Wallis, 2004; Galbraith and van Norden, 2012). The availability of a "benchmark" Bayesian analysis is thus of particular importance for practitioners.

Firstly, we consider univariate (continuous) two-piece distributions with different scales on both sides of the location parameter. Then, we focus on the family of reparameterizations defined in Arellano-Valle et al. (2005), where the scales are reparameterized in terms of a common scale and a skewness parameter. Whereas we discuss orthogonality of parameterizations, which is of direct interest for likelihood-based frequentist inference, we will mostly focus on Bayesian inference in this paper. A commonly used prior structure to reflect an absence of prior information is the Jeffreys (or "Jeffreys-rule") prior, which is the reference prior (Berger et al., 2009) in the case of a scalar parameter under asymptotic posterior normality. Under these conditions, Clarke and Barron (1994) showed that this prior asymptotically maximizes the expected information from repeated sampling. The Jeffreys prior is an interesting choice because no subjective parameters have to be elicited and it is invariant under reparameterizations (Jeffreys, 1941; Ibrahim and Laud, 1991).

However, in our two-piece location-scale framework (and its reparameterizations), we show that Jeffreys prior does not lead to a proper posterior in the wide and empirically interesting class of distributions obtained by skewing scale mixtures of normals. In addition, we consider the independence Jeffreys prior (constructed as the product of the Jeffreys priors for each parameter while considering the other parameters are fixed), which is shown to lead to a proper posterior under some parameterizations. Simple conditions regarding posterior existence with the independence Jeffreys prior are derived. We propose an alternative prior structure, which is partly subjective, but which is easily elicited and leads to valid Bayesian inference in a wide and practically relevant class of parameterizations of two-piece models.

The structure of this document is as follows: in Section 2 we present the two-piece location-scale model and the family of parameterizations defined in Arellano-Valle et al. (2005). We derive the Fisher information matrix for these models as well as the Jeffreys and independence Jeffreys priors. In Section 3 we examine posterior existence with these priors in the context of a scale mixture of normals for the underlying symmetric distribution. We also propose two alternative prior structures, one of which is our recommended prior choice for users of these models. In Section 4 we present an application of the Bayesian models studied here on a real data set. The final section contains concluding remarks. Proofs of all theorems as well as a numerical coverage analysis of the 95% credible intervals for various models are given in the supplementary material.

2 Sampling Models and Jeffreys priors

2.1 Two-piece location-scale models

Let $f(y|\mu, \sigma)$ be an absolutely continuous density with support on $\mathbb{R}$, location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma \in \mathbb{R}^+$, and denote $f\left(\frac{y-\mu}{\sigma}|0,1\right) = f\left(\frac{y-\mu}{\sigma}\right)$. Consider the following "two-piece" density constructed of $f\left(\frac{y-\mu}{\sigma_1}\right)$ truncated to $(-\infty, \mu)$ and $f\left(\frac{y-\mu}{\sigma_2}\right)$ truncated to $[\mu, \infty)$:

$$g(y|\mu, \sigma_1, \sigma_2, \varepsilon) = \frac{2\varepsilon}{\sigma_1} f\left(\frac{y-\mu}{\sigma_1}\right) I_{(-\infty,\mu)}(y) + \frac{2(1-\varepsilon)}{\sigma_2} f\left(\frac{y-\mu}{\sigma_2}\right) I_{[\mu,\infty)}(y), \tag{1}$$

where $\sigma_1 \in \mathbb{R}^+$ and $\sigma_2 \in \mathbb{R}^+$ are separate scale parameters and $0 < \varepsilon < 1$. To get a continuous density, we need to choose $\varepsilon = \sigma_1/(\sigma_1 + \sigma_2)$, so that

$$s(y|\mu, \sigma_1, \sigma_2) = \frac{2}{\sigma_1 + \sigma_2} \left[f\left(\frac{y-\mu}{\sigma_1}\right) I_{(-\infty,\mu)}(y) + f\left(\frac{y-\mu}{\sigma_2}\right) I_{[\mu,\infty)}(y) \right]. \tag{2}$$

Typically, f will be a symmetric density function. In this paper, we will assume f to be symmetric with a single mode at zero, which means that μ is the mode of the density in (2). If we choose f to be a normal or a Student t density, the distribution in (2) corresponds to split-normal and split-t distributions, respectively, as defined in Geweke (1989). In earlier work, the case with normal f was termed joined half-Gaussian by Gibbons and Mylroie (1973) and two-piece normal by John (1982). A historical account of the many guises of this distribution is provided in Wallis (2013). In line with most of the recent literature (Jones, 2006; Jones and Anaya-Izquierdo, 2011; Wallis, 2013), we shall denote the model in (2) as the two-piece model. Since

$$\int_{-\infty}^{\mu} s(y|\mu, \sigma_1, \sigma_2)\, dy = \frac{\sigma_1}{\sigma_1 + \sigma_2}, \tag{3}$$

s is skewed about μ if $\sigma_1 \neq \sigma_2$ and the ratio σ_1/σ_2 controls the allocation of mass to each side of μ.

We are mainly interested in the inferential properties of these skewed distributions under the popular Jeffreys priors, but will also briefly discuss orthogonality of their parameters. Cox and Reid (1987) define two parameters, θ_1 and θ_2, to be orthogonal if the corresponding off-diagonal entry of the Fisher information matrix is zero. If θ_1 is orthogonal to θ_2, we will denote this as $\theta_1 \perp \theta_2$.

We first calculate the Fisher information matrix and characterize, through the symmetric density f, the cases where this matrix is well defined:

Theorem 1 *Let $s(y|\mu, \sigma_1, \sigma_2)$ be as in (2) and suppose that the following conditions hold*

(i) $\int_0^\infty \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt < \infty$,

(ii) $\int_0^\infty t^2 \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt < \infty$,

(iii) $\lim_{t\to\infty} t f(t) = 0$ *or* $\int_0^\infty t f'(t) dt = -\frac{1}{2}$*, which means that* $f(t)$ *is* $o\left(\frac{1}{t}\right)$.

Then the Fisher information matrix $I(\mu, \sigma_1, \sigma_2)$ *is*

$$\begin{pmatrix} \frac{2\alpha_1}{\sigma_1\sigma_2} & -\frac{2\alpha_3}{\sigma_1(\sigma_1+\sigma_2)} & \frac{2\alpha_3}{\sigma_2(\sigma_1+\sigma_2)} \\ -\frac{2\alpha_3}{\sigma_1(\sigma_1+\sigma_2)} & \frac{\alpha_2}{\sigma_1(\sigma_1+\sigma_2)} + \frac{\sigma_2}{\sigma_1(\sigma_1+\sigma_2)^2} & -\frac{1}{(\sigma_1+\sigma_2)^2} \\ \frac{2\alpha_3}{\sigma_2(\sigma_1+\sigma_2)} & -\frac{1}{(\sigma_1+\sigma_2)^2} & \frac{\alpha_2}{\sigma_2(\sigma_1+\sigma_2)} + \frac{\sigma_1}{\sigma_2(\sigma_1+\sigma_2)^2} \end{pmatrix}, \tag{4}$$

where

$$\begin{aligned} \alpha_1 &= \int_0^\infty \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt, \\ \alpha_2 &= 2\int_0^\infty \left[1 + t\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt = -1 + 2\int_0^\infty t^2 \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt, \\ \alpha_3 &= \int_0^\infty t \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt. \end{aligned}$$

Conditions (i) and (ii) are required for the existence of the expression in (4) and are satisfied under regularity conditions (Lehmann and Casella, 1998; p. 126). Condition (iii) is useful to simplify some expressions and is satisfied by many models of interest. As examples, normal, Student t, logistic, Cauchy, Laplace and exponential power distributions (Box and Tiao, 1973; p. 157) all satisfy $(i) - (iii)$. Given that α_1, α_2 and α_3 are positive as long as

$f'(t) \neq 0$ everywhere, none of the entries of the Fisher information matrix are zero. Therefore, this is a non-orthogonal parameterization.

The Jeffreys prior, proposed by Jeffreys (1941), is defined as the square root of the determinant of the Fisher information matrix. In contrast, the independence Jeffreys prior is defined as the product of the Jeffreys priors for each parameter independently, while treating the others parameters as fixed.

Corollary 1 *If the Fisher information matrix in (4) is non-singular, then the Jeffreys prior for the parameters in (2) is*

$$\pi_J(\mu, \sigma_1, \sigma_2) \quad \propto \quad \frac{1}{\sigma_1\sigma_2(\sigma_1 + \sigma_2)}. \tag{5}$$

The independence Jeffreys prior is

$$\pi_I(\mu, \sigma_1, \sigma_2) \quad \propto \quad \frac{\sqrt{[\sigma_1 + \alpha_2(\sigma_1 + \sigma_2)][\sigma_2 + \alpha_2(\sigma_1 + \sigma_2)]}}{\sqrt{\sigma_1\sigma_2}(\sigma_1 + \sigma_2)^2}. \tag{6}$$

The Jeffreys prior is defined only in the cases when the Fisher information matrix is non-singular. The determinant of the Fisher information matrix can be factored into two terms, one dependent on the parameters and the other dependent on the constants $(\alpha_1, \alpha_2, \alpha_3)$. The former is always positive. The following result gives conditions on the density f that ensure that the second factor does not vanish and the Fisher information matrix is thus non-singular.

Theorem 2 *If the conditions of Theorem* 1 *are satisfied and* $f'(t) \neq 0$ $a.e.$, *then the Fisher information matrix is non-singular.*

In particular, the Fisher information matrix (4) is non-singular if f corresponds to a normal, Laplace, exponential power, logistic, Cauchy or Student t distribution. The structure of the independence Jeffreys prior in (6) assumes that $\alpha_2 > 0$, which will always be the case (see the proof of Theorem 2 in the supplementary material).

2.2 Reparameterizations of the two-piece model

To link the two-piece model in (2) with the family defined in Arellano-Valle et al. (2005), we use the following reparameterization (one-to-one transformation)

$$\begin{aligned} (\mu, \sigma_1, \sigma_2) &\leftrightarrow (\mu, \sigma, \gamma), \\ \mu &= \mu, \\ \sigma_1 &= \sigma b(\gamma), \\ \sigma_2 &= \sigma a(\gamma), \end{aligned} \tag{7}$$

where $\gamma \in \Gamma$, $\sigma > 0$ and $a(\gamma) > 0$ and $b(\gamma) > 0$ are differentiable functions such that

$$0 < |\lambda(\gamma)| < \infty, \text{with } \lambda(\gamma) \equiv \frac{d}{d\gamma} \log \left[\frac{a(\gamma)}{b(\gamma)}\right]. \tag{8}$$

The condition in (8) implies that (7) is a non-singular mapping and is thus necessary for it to be a one-to-one transformation. Then we get the following reparameterized density from (2)

$$s(y|\mu, \sigma, \gamma) = \frac{2}{\sigma[a(\gamma) + b(\gamma)]} \left[f\left(\frac{y-\mu}{\sigma b(\gamma)}\right) I_{(-\infty,\mu)}(y) + f\left(\frac{y-\mu}{\sigma a(\gamma)}\right) I_{[\mu,\infty)}(y) \right]. \tag{9}$$

This expression was presented by Arellano-Valle et al. (2005) as a general class of asymmetric distributions, which includes various skewed distributions presented in the literature. Like Jones (2006), we view (9) with a given choice of f not as a class of densities but as a class of reparameterizations of the same density.

Two parameterizations using the functions $\{a(\gamma), b(\gamma)\}$ have been widely studied: the inverse scale factors (ISF) model (Fernández and Steel, 1998), corresponding to $\{a(\gamma), b(\gamma)\} = \{\gamma, 1/\gamma\}$ for $\gamma \in \mathbb{R}^+$ and the ϵ-skew model (Mudholkar and Hutson, 2000), which chooses $\{a(\gamma), b(\gamma)\} = \{1 + \gamma, 1 - \gamma\}$ for $\gamma \in (-1, 1)$.

The Fisher information matrix for the reparameterized model in (9) is:

Theorem 3 *Let $f(y|\mu,\sigma)$ be as in Theorem 1. Then the Fisher information matrix $I(\mu,\sigma,\gamma)$ for model (9) is*

$$\begin{pmatrix} \frac{2\alpha_1}{a(\gamma)b(\gamma)\sigma^2} & 0 & \frac{2\alpha_3}{\sigma[a(\gamma)+b(\gamma)]}\left[\frac{a'(\gamma)}{a(\gamma)}-\frac{b'(\gamma)}{b(\gamma)}\right] \\ 0 & \frac{\alpha_2}{\sigma^2} & \frac{\alpha_2}{\sigma}\left[\frac{a'(\gamma)+b'(\gamma)}{a(\gamma)+b(\gamma)}\right] \\ \frac{2\alpha_3}{\sigma[a(\gamma)+b(\gamma)]}\left[\frac{a'(\gamma)}{a(\gamma)}-\frac{b'(\gamma)}{b(\gamma)}\right] & \frac{\alpha_2}{\sigma}\left[\frac{a'(\gamma)+b'(\gamma)}{a(\gamma)+b(\gamma)}\right] & \frac{\alpha_2+1}{a(\gamma)+b(\gamma)}\left[\frac{b'(\gamma)^2}{b(\gamma)}+\frac{a'(\gamma)^2}{a(\gamma)}\right]-\left[\frac{a'(\gamma)+b'(\gamma)}{a(\gamma)+b(\gamma)}\right]^2 \end{pmatrix}.$$

The fact that the elements I_{12} and I_{21} are zero indicates that this reparameterization is interesting because it induces orthogonality between the parameters μ and σ for any choice of $\{a(\gamma),b(\gamma)\}$. In addition, by appropriately choosing the pair of functions $\{a(\gamma),b(\gamma)\}$ we can generate more zero entries in the Fisher information matrix, as shown in the following corollary.

Corollary 2 *If $\frac{d}{d\gamma}\log\left[a(\gamma)+b(\gamma)\right]=0$, then $I_{23}=I_{32}=0$. In particular if $a(\gamma)+b(\gamma)$ is constant, then $I_{23}=I_{32}=0$.*

If $\alpha_3>0$, then $I_{13}=I_{31}=0$ only if $a(\gamma)\propto b(\gamma)$ which does not satisfy (8). Jones and Anaya-Izquierdo (2011) analysed the zeroes of the expectation of the Hessian matrix of (μ,σ,γ) in model (9) augmented with an extra parameter to model the properties of f. They also found that $\mu\perp\sigma$ and if $a(\gamma)+b(\gamma)$ is constant then $\sigma\perp\gamma$ as in Corollary 2.

The corresponding Jeffreys prior and independence Jeffreys prior for the parameterization (μ,σ,γ) are given in the following result.

Corollary 3 *If the Fisher information matrix is non-singular, then the Jeffreys prior for the parameters in (9) is*

$$\pi_J(\mu,\sigma,\gamma)\propto\frac{|a'(\gamma)b(\gamma)-a(\gamma)b'(\gamma)|}{\sigma^2a(\gamma)b(\gamma)[a(\gamma)+b(\gamma)]}=\frac{|\lambda(\gamma)|}{\sigma^2[a(\gamma)+b(\gamma)]},\qquad(10)$$

where $\lambda(\gamma)$ was defined in (8). The independence Jeffreys prior is

$$\pi_I(\mu,\sigma,\gamma)\quad\propto\quad\frac{1}{\sigma}\sqrt{\frac{\alpha_2+1}{a(\gamma)+b(\gamma)}\left[\frac{b'(\gamma)^2}{b(\gamma)}+\frac{a'(\gamma)^2}{a(\gamma)}\right]-\left[\frac{a'(\gamma)+b'(\gamma)}{a(\gamma)+b(\gamma)}\right]^2}.\qquad(11)$$

Conditions to ensure non-singularity of the Fisher information matrix for the parameterization in (9) are similar to those obtained for the two-piece model (2) in Theorem 2. The only difference is that in this case we have to choose a pair of functions $\{a(\gamma), b(\gamma)\}$ such that (7) corresponds to a non-singular transformation:

Corollary 4 *If the conditions of Theorem 1 are satisfied, $f'(t) \neq 0$ a.e., and (8) holds, then the Fisher information matrix corresponding to model (9) is non-singular.*

Due to the invariance property of the Jeffreys prior there is a one-to-one relationship between (5) and (10). On the other hand, the independence Jeffreys prior is not invariant under reparameterizations, so the properties of this prior are dependent on the choice of $\{a(\gamma), b(\gamma)\}$.

Now we will briefly discuss the inverse scale factors and ϵ-skew models.

2.2.1 Inverse scale factors model

The ISF model corresponds to choosing $\{a(\gamma) = \gamma, b(\gamma) = 1/\gamma\}$, $\gamma \in \mathbb{R}^+$ in (9), so that from Theorem 3 the Fisher information matrix of the parameters (μ, σ, γ) is

$$I(\mu, \sigma, \gamma) = \begin{pmatrix} \frac{2\alpha_1}{\sigma^2} & 0 & \frac{4\alpha_3}{\sigma(\gamma^2+1)} \\ 0 & \frac{\alpha_2}{\sigma^2} & \frac{\alpha_2\left(\gamma^2-1\right)}{\sigma(\gamma^3+\gamma)} \\ \frac{4\alpha_3}{\sigma(\gamma^2+1)} & \frac{\alpha_2\left(\gamma^2-1\right)}{\sigma(\gamma^3+\gamma)} & \frac{\alpha_2}{\gamma^2} + \frac{4}{(\gamma^2+1)^2} \end{pmatrix}. \tag{12}$$

If the Fisher information matrix in (12) is non-singular, then the Jeffreys prior for the ISF model is

$$\pi_J(\mu, \sigma, \gamma) \quad \propto \quad \frac{1}{\sigma^2\left(1+\gamma^2\right)}, \tag{13}$$

which has a finite integral over $\gamma \in \mathbb{R}^+$, but is improper in μ and σ. The independence Jeffreys prior is

$$\pi_I(\mu, \sigma, \gamma) \quad \propto \quad \frac{1}{\sigma}\sqrt{\frac{\alpha_2}{\gamma^2} + \frac{4}{\left(\gamma^2+1\right)^2}}, \tag{14}$$

which is not integrable in any of the parameters.

2.2.2 ϵ-skew model

For the ϵ-skew model we choose $\{a(\gamma) = 1 - \gamma, b(\gamma) = 1 + \gamma\}$ in (9), where $\gamma \in (-1, 1)$, leading to the Fisher information matrix

$$I(\mu, \sigma, \gamma) = \begin{pmatrix} \frac{2\alpha_1}{\sigma^2(1-\gamma^2)} & 0 & -\frac{2\alpha_3}{\sigma(1-\gamma^2)} \\ 0 & \frac{\alpha_2}{\sigma^2} & 0 \\ -\frac{2\alpha_3}{\sigma(1-\gamma^2)} & 0 & \frac{\alpha_2+1}{1-\gamma^2} \end{pmatrix}. \tag{15}$$

The ϵ-skew parameterization satisfies the condition in Corollary 2 and thus its Fisher information matrix has four zeroes. The presence of zero entries often simplifies classical inference (Jones and Anaya-Izquierdo, 2011). For example, in the cases where f is normal or Laplace, the corresponding ϵ-skew model leads to maximum likelihood estimators in closed form (Mudholkar and Hutson, 2000; Arellano-Valle et al., 2005).

Provided the Fisher information matrix in (15) is non-singular, the Jeffreys prior for the ϵ-skew model is

$$\pi_J(\mu, \sigma, \gamma) \quad \propto \quad \frac{1}{\sigma^2(1-\gamma^2)}, \tag{16}$$

which is not integrable in any of the parameters. The independence Jeffreys prior is

$$\pi_I(\mu, \sigma, \gamma) \quad \propto \quad \frac{1}{\sigma\sqrt{1-\gamma^2}}, \tag{17}$$

which has a finite integral over $\gamma \in (-1, 1)$, but does not integrate in μ and σ. For this model the independence Jeffreys prior does not depend on f (through α_2), in contrast with the priors for the two-piece model in (6) and the ISF model in (14).

In the different models mentioned above, the skewness parameter γ does not have the same interpretation. This makes it particularly difficult to compare models and priors on γ. However, they can be compared by using a skewness measure that has the same interpretation across parameterizations. Here we use the skewness measure with respect to the mode from Arnold and Groeneveld (1995), defined as

Definition 1 *The Arnold-Groeneveld measure of skewness for a distribution function S corresponding to a unimodal density with the mode at M is defined as 1 minus twice the probability mass to the left of the mode:*

$$AG = 1 - 2S(M).$$

The AG measure takes values in $(-1, 1)$ and can be interpreted as the difference between the mass to the right and the mass to the left of the mode. Positive values of AG indicate right skewness while negative values indicate left skewness. From (3) it is immediate that for the two-piece model $AG = (\sigma_2 - \sigma_1)/(\sigma_1 + \sigma_2)$, which only depends on the two scales and not on the properties of f. Similarly, for the parameterization in Arellano-Valle et al. (2005) in (9) the AG skewness measure has a closed form which only depends on γ:

$$AG(\gamma) = \frac{a(\gamma) - b(\gamma)}{a(\gamma) + b(\gamma)}.$$

For the special case of the ISF model in Subsection 2.2.1, this reduces to

$$AG(\gamma) = \frac{\gamma^2 - 1}{\gamma^2 + 1},$$

while for the ϵ-skew model in Subsection 2.2.2 we obtain $AG(\gamma) = -\gamma$.

In both examples above, the AG skewness measure is a monotonic function of γ, so we can meaningfully interpret γ as a skewness parameter. In general, we will be mostly interested in parameterizations where AG is a monotonic function of γ, which can be characterized as follows:

Theorem 4 *Let s, $a(\gamma)$ and $b(\gamma)$ be as in (9), then for any unimodal density f*

- *$AG(\gamma)$ is increasing if and only if $\lambda(\gamma) > 0$.*
- *$AG(\gamma)$ is decreasing if and only if $\lambda(\gamma) < 0$.*

3 Inference

In this section we will present necessary and/or sufficient conditions for the properness of the posterior distribution of the parameters of the two-piece models considered when using the priors presented in the previous section, as well as two alternative priors to be introduced later in Subsection 3.4. Throughout this section we will assume that we have observed a sample of n independent replications from either (2) or (9). Although those models are equivalent up to a reparameterization, we will show that the existence of the posterior distribution can depend on the parameterization, if the prior is not invariant under reparameterization. We separately deal with samples where all the observations are different and samples which contain repeated observations. Most of the results in this section are for the case where the underlying symmetric distribution (with density f) belongs to the wide class of scale mixtures of normals. Of course, a meaningful use of the results in Subsections 3.1 and 3.2 implies a nonsingular information matrix (see Theorem 2 and Corollary 4) so that the Jeffreys prior exists or implies that the independence Jeffreys prior is well-defined. However, most cases of practical interest will correspond to an f that allows for these priors to be well-defined.

Recall that a density f corresponds to a scale mixture of normals if it can be written as

$$f(x) = \int_0^\infty \tau^{1/2}\phi(\tau^{1/2}x)dP_\tau,$$

where ϕ is the standard normal density and P_τ is a mixing distribution on $\mathbb{R}_+$. The class of scale mixtures of normals is quite a rich class of symmetric and unimodal continuous distributions and contains many popular distributions, such as the normal, Student t, logistic, Laplace, Cauchy and the exponential power family with power $1 \le q < 2$ (see Fernández and Steel, 2000 for more details). This class does not cover distributions with tails thinner than normal tails.

3.1 Independence Jeffreys prior

The independence Jeffreys prior is not invariant under reparameterizations. Therefore if we consider one-to-one transformations as in (7), we need to analyse the properness of the posterior distribution of (μ, σ, γ) for each specific choice of $\{a(\gamma), b(\gamma)\}$. Thus, we examine the models in (2) and (9) separately.

Theorem 5 *Let $\mathbf{y} = (y_1, \dots, y_n)$ be an independent sample from the model in (2), where f is a scale mixture of normals. Then,*

(i) The posterior distribution of $(\mu, \sigma_1, \sigma_2)$ using the independence Jeffreys prior (6) is proper if $n \geq 2$ and all the observations are different.

(ii) Suppose that the sample $\mathbf{y}$ contains repeated observations. Let k be the largest number of observations with the same value in $\mathbf{y}$. If $1 < k < n$, then the posterior of $(\mu, \sigma_1, \sigma_2)$ is proper if and only if the mixing distribution of f satisfies

$$\int_{0<\tau_1\leq\cdots\leq\tau_n<\infty} \tau_{n-k}^{-(n-2)/2} \prod_{i\neq n-k,n} \tau_i^{1/2} dP_{(\tau_1,\dots,\tau_n)} < \infty, \tag{18}$$

where $dP_{(\tau_1,\dots,\tau_n)}$ denotes the distribution of the n mixing parameters τ_j, $j = 1, \dots, n$, associated with the n observations. In the case of the two-piece normal sampling model (i.e. normal f), it suffices to have two different observations.

Thus, for a wide and practically important class of distributions f, the two-piece model in (2) with the independence Jeffreys prior leads to valid inference in (almost) any sample of two or more observations. Equation (18) establishes a condition on the tails of the mixing distribution that leads to a proper posterior distribution using the independence Jeffreys prior. We refer the reader to Fernández and Steel (1999) for more details on this condition.

For the model in (9), we can derive useful existence results within a class of prior distributions:

Theorem 6 *Let* $\mathbf{y} = (y_1, \ldots, y_n)$ *be an independent sample from the model in* (9)*, where* f *is a scale mixture of normals. Consider a prior distribution of the form* $\pi(\mu, \sigma, \gamma) \propto \sigma^{-1}\pi(\gamma)$*, for some* $\pi(\gamma)$*. Then:*

(i) a necessary condition for the properness of the posterior distribution of (μ, σ, γ) *is*

$$\int_\Gamma \left[\frac{a(\gamma)}{a(\gamma) + b(\gamma)}\right]^n \pi(\gamma)\, d\gamma < \infty. \tag{19}$$

(ii) the posterior distribution of (μ, σ, γ) *is proper if* $n \geq 2$*, all the observations are different, and* $\pi(\gamma)$ *is proper.*

(iii) Suppose that the sample $\mathbf{y}$ *contains repeated observations and* $\pi(\gamma)$ *is proper. Let* k *be the largest number of observations with the same value in* $\mathbf{y}$*. If* $1 < k < n$*, then the posterior of* (μ, σ, γ) *is proper if and only if the mixing distribution of* f *satisfies* (18)*. In the case of the two-piece normal sampling model (i.e. normal* f*), it suffices to have two different observations.*

This theorem implies that a posterior will exist for the ϵ-skew model under the independence Jeffreys prior in (17), as this prior is a member of the class in Theorem 6 with proper $\pi(\gamma)$.

However, for the ISF model the independence Jeffreys prior does not integrate in γ and we can show that the necessary condition (19) is violated, so that a posterior does not exist in this case:

Corollary 5 *If* f *is a scale mixture of normals in (9) and* $\{a(\gamma), b(\gamma)\}$ *are as in the inverse scale factors model, then the posterior distribution of* (μ, σ, γ) *is improper under the independence Jeffreys prior (14).*

Theorem 6 emphasizes the relevance of the choice of the functions $\{a(\gamma), b(\gamma)\}$ for the properness of the posterior distribution of (μ, σ, γ) when using the independence Jeffreys prior.

In particular, condition (19) can be used to detect parameterizations $\{a(\gamma), b(\gamma)\}$ that produce improper posteriors. The fact that the ISF model does not allow for inference with the independence Jeffreys prior is rather surprising since this prior almost always leads to proper posteriors, and the ISF model is quite a straightforward extension of the usual location-scale model. Subsection 3.3 will shed more light on this.

3.2 Jeffreys prior

We now examine the properness of the posterior distribution of the parameters (μ, σ, γ) under the Jeffreys prior. An important feature of this prior is the invariance under one-to-one reparameterizations. Therefore, the results regarding the properness of the posterior of (μ, σ, γ) for any choice of $\{a(\gamma), b(\gamma)\}$ in model (9) that corresponds to a one-to-one transformation in (7) are the same and also applicable to the posterior of $(\mu, \sigma_1, \sigma_2)$ in model (2).

Theorem 7 *Let s be as in (9), assume that f is a scale mixture of normals and consider the Jeffreys prior (10) for the parameters of this model. Then, for $n \geq 2$, a necessary condition for the properness of the posterior distribution of (μ, σ, γ) is*

$$\int_{\Gamma} \left[\frac{a(\gamma)}{a(\gamma)+b(\gamma)}\right]^{n+1} |\lambda(\gamma)|\, d\gamma < \infty, \tag{20}$$

with $\lambda(\gamma)$ defined as in (8).

Corollary 6 *Consider sampling from (9) with f a scale mixture of normals and $\{a(\gamma), b(\gamma)\}$ as in the inverse scale factors model, then the posterior distribution of (μ, σ, γ) is improper using the Jeffreys prior (10). As a consequence, for any pair of functions $\{a(\gamma), b(\gamma)\}$ such that the mapping $(\mu, \sigma_1, \sigma_2) \leftrightarrow (\mu, \sigma, \gamma)$ is one-to-one, the posterior distribution of (μ, σ, γ) is improper using the Jeffreys prior (10).*

Proof. *We can verify that the necessary condition (20) is not satisfied for these functions.*

This corollary implies that we can not conduct Bayesian inference for the parameters of this type of skewed distributions using the Jeffreys prior. It is rather rare to find that the

Jeffreys prior does not lead to a proper posterior, and it is somewhat surprising to find that we can not use this prior in these rather simple classes of two-piece distributions with only three parameters.

Because the Jeffreys prior is invariant to reparameterization, its use is thus prohibited in any one-to-one reparameterization of the two-piece models in (2) or (9). However, one way to get around this problem is to choose functions $\{a(\gamma), b(\gamma)\}$ such that the mapping $(\mu, \sigma, \gamma) \mapsto (\mu, \sigma_1, \sigma_2)$ is not one-to-one, but hopefully still of some interest for modelling. Another way to produce a proper posterior distribution when using the Jeffreys prior is to restrict Γ such that $\lambda(\gamma)$ is absolutely integrable.

Theorem 8 *Let s be as in (9) where f is normal or Laplace. Consider the Jeffreys prior (10) for the parameters of this model. Let $\{a(\gamma), b(\gamma)\}$ be continuously differentiable functions for $\gamma \in \Gamma$ such that*

$$0 < \int_\Gamma |\lambda(\gamma)| \, d\gamma < \infty. \tag{21}$$

Then we have the following results

(i) The posterior distribution of (μ, σ, γ) is proper when $n \geq 2$ and there are at least two different observations.

(ii) The mapping $(\mu, \sigma, \gamma) \mapsto (\mu, \sigma_1, \sigma_2)$ is not one-to-one.

(iii) If Γ is an interval (not necessarily bounded) and $AG(\gamma)$ is monotonic, then $AG(\gamma)$ is not surjective.

First, we considered forcing existence of the posterior through the choice of the functions $\{a(\gamma), b(\gamma)\}$, in particular such that the ratio $a(\gamma)/b(\gamma)$ is bounded, which excludes a one-to-one reparameterization in (7). However, the examples we generated in this way did not lead to implied priors on AG that could be of interest to practitioners.

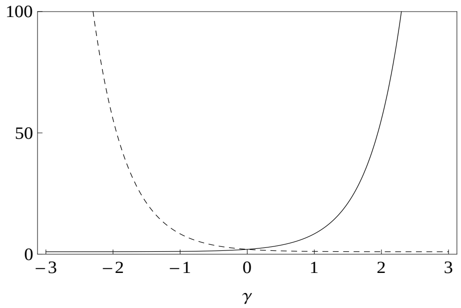
It is actually easier to generate examples of practical relevance if we restrict the parameter space of γ in the context of functions $\{a(\gamma), b(\gamma)\}$ that would not lead to a proper posterior with unrestricted γ. The following is such an example.

Example 1 (Logistic AG) *Consider $a(\gamma) = 1 + \exp(2\gamma)$, $b(\gamma) = 1 + \exp(-2\gamma)$ for $\gamma \in \mathbb{R}$, then*

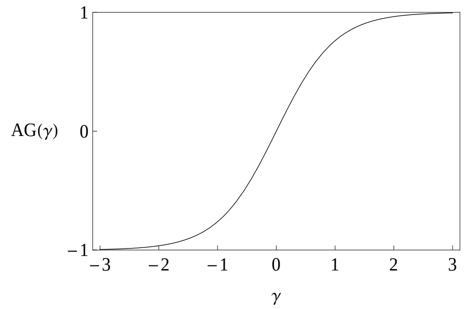
$$\begin{aligned} AG(\gamma) &= \tanh(\gamma), \\ \lambda(\gamma) &= 2, \\ \pi_J(\mu, \sigma, \gamma) &\propto \frac{1}{\sigma^2} \operatorname{sech}(\gamma)^2, \end{aligned} \tag{22}$$

where $\tanh(\cdot)$ and $\operatorname{sech}(\cdot)$ denote the hyperbolic tangent and the hyperbolic secant functions. In addition, the functions $a(\gamma)$, $b(\gamma)$ and $AG(\gamma)$ are monotonic $\forall\ \gamma \in \mathbb{R}$, the Jeffreys prior in (22) implies that $AG \sim \text{Unif}(-1, 1)$ and $AG : \mathbb{R} \mapsto (-1, 1)$. Clearly, $\lambda(\gamma)$ is not integrable on $\mathbb{R}$, but if we restrict $\gamma \in [-B, B]$ for some $0 < B < \infty$, then we can use the Jeffreys prior (22) for making inference on (μ, σ, γ) for normal or Laplace f and $AG : \mathbb{R} \mapsto [\tanh(-B), \tanh(B)]$. Figure 1 presents the functions $a(\gamma)$, $b(\gamma)$ and $AG(\gamma)$ for $B = 3$. The induced prior on AG is a Uniform over the range $[\tanh(-B), \tanh(B)] = [-0.995, 0.995]$.

We will call the model in Example 1 the "logistic AG model" as $AG(\gamma)$ is a logistic function of γ transformed to take values in the interval (-1,1) for $\gamma \in \mathbb{R}$. The choice of $a(\gamma)$ and $b(\gamma)$ does lead to a one-to-one transformation in (7) when $\gamma \in \mathbb{R}$, but not if γ is restricted to a bounded interval: then the ratio $a(\gamma)/b(\gamma)$ is also bounded and this precludes a one-to-one mapping. $a(\gamma)$ and $b(\gamma)$ satisfy the condition $a(\gamma)+b(\gamma) = a(\gamma)b(\gamma)$, which induces a uniform prior on the skewness measure $AG(\gamma)$. This might be an attractive prior for practitioners to use in the absence of strong prior information.



(a)



(b)

Figure 1: (a) $a(\gamma)$ (solid line) and $b(\gamma)$ (dashed line); (b) $AG(\gamma)$.

3.3 Intuitive explanation

The lack of existence of a posterior distribution under a commonly used prior in what is essentially a very simply generalisation of a standard location-scale model can be considered surprising. Thus, we offer a few explanatory comments in this subsection. These are not meant to be formal proofs (they can be found in the supplementary material), but merely intuitive ideas that help us understand what drives the main results we have found in the previous subsections.

In the context of the two-piece model in (2), it is easy to see that as σ_1 tends to zero, the sampling density tends to the half density on $[\mu, \infty)$ with scale σ_2. Thus, the likelihood will be constant in σ_1 in the neighbourhood of zero. This means the prior needs to integrate in that neighbourhood for a posterior to exist. If we consider the independent Jeffreys prior in (6) it behaves like $\sigma_1^{-1/2}$ for small σ_1 and this integrates close to zero. Indeed, we have a posterior in this case. However, the Jeffreys prior in (5) behaves like $1/\sigma_1$ for small σ_1 and this does not integrate, thus precluding a posterior. Of course, similar arguments hold for small σ_2.

In the case of the reparameterized model in (9), we have a potential problem if one of the scales, say, $\sigma a(\gamma)$ goes to zero. If then the ratio $b(\gamma)/a(\gamma)$ has an upper bound, this will necessarily imply that both scales tend to zero, so the model behaves like a standard location-scale model which leads to a proper posterior under the Jeffreys prior. This is the case explored in Theorem 8 and Example 1. If, however, the ratio between the functions $a(\gamma)$ and $b(\gamma)$ is not bounded and (7) defines a one-to-one mapping, we will have no posterior with the Jeffreys prior due to the invariance of this prior under reparameterization, and it depends on the particular choice of functions $\{a(\gamma), b(\gamma)\}$ whether the independence Jeffreys prior will lead to a posterior. It is helpful to transform the parameters back to those of the two-piece model in (2). Then, for the ϵ-skew model the independence Jeffreys prior in (17) can be shown to behave like $\sigma_i^{-1/2}$ for small $\sigma_i, i = 1, 2$, which is integrable close to zero, and the posterior is well-defined. On the other hand, the independence Jeffreys prior for the ISF model in (14) behaves like $1/\sigma_i$ for small $\sigma_i, i = 1, 2$, which does not integrate in a neighbourhood of zero and precludes posterior existence.

3.4 Alternative priors

We now introduce two alternative priors for the sampling model in (9): one is a modification of the Jeffreys prior and the other is a non-objective prior with an elicitation strategy through an easily interpretable quantity and the possibility to use vague priors. Both prior structures will be of the form

$$\pi(\mu, \sigma, \gamma) \propto \sigma^{-1}\pi(\gamma). \tag{23}$$

3.4.1 Modified Jeffreys prior

The first choice for $\pi(\gamma)$ in (23) consists of the factor dependent on γ of the Jeffreys prior (10), which implies

$$\begin{aligned} \pi_M(\gamma) &\propto \frac{|a'(\gamma)b(\gamma) - a(\gamma)b'(\gamma)|}{a(\gamma)b(\gamma)[a(\gamma) + b(\gamma)]} \\ &= \frac{1}{a(\gamma) + b(\gamma)} \left| \frac{d}{d\gamma} \log \left[\frac{a(\gamma)}{b(\gamma)} \right] \right|. \end{aligned} \tag{24}$$

The resulting modified Jeffreys prior can also be interpreted as the independence Jeffreys prior with the independence applied to the two blocks μ and (σ, γ), rather than the three parameters separately (see Fonseca et al., 2008 for a similar prior in the context of a Student-t regression model with unknown degrees of freedom).

3.4.2 AG beta prior

The second alternative prior $\pi_\beta(\gamma)$ is such that $\delta = (AG+1)/2$, the AG skewness measure rescaled to the unit interval, has a Beta(α_0, β_0) distribution. Thus, this prior is not obtained through a formal rule and can be elicited on the basis of AG, which has a clear interpretation as the difference between the mass to the right and the mass to the left of the mode (see Definition 1). In practice, this prior is perhaps most useful for values of α_0 and β_0 relatively close to one, reflecting vague prior information on the AG measure of skewness. For γ it corresponds to

$$\pi_\beta(\gamma) \propto \frac{|a'(\gamma)b(\gamma) - a(\gamma)b'(\gamma)|}{[a(\gamma)+b(\gamma)]^{\alpha_0+\beta_0}} a(\gamma)^{\alpha_0-1} b(\gamma)^{\beta_0-1}. \tag{25}$$

Despite being motivated in rather different ways, both alternative priors coincide in certain special cases. In particular, prior (24) implies that $\delta \sim$ Beta$(1/2, 1/2)$ if $a(\gamma)b(\gamma) = c$. This is the case of the Inverse Scale Factors parameterization. In addition, the prior distributions (24) and (25) coincide if $\alpha_0 = \beta_0 = 1$ and $a(\gamma) + b(\gamma) = a(\gamma)b(\gamma)$, as already remarked in the context of the logistic AG model in Example 1.

The alternative priors of (μ, σ, γ) for the Inverse Scale Factors model are respectively

$$\pi_M(\mu, \sigma, \gamma) \propto \frac{1}{\sigma\,(1+\gamma^2)}, \tag{26}$$

$$\pi_\beta(\mu, \sigma, \gamma) \propto \frac{\gamma^{2\alpha_0-1}}{\sigma\,(1+\gamma^2)^{\alpha_0+\beta_0}}, \tag{27}$$

for $\gamma \in \mathbb{R}^+$. Indeed both priors coincide when $\alpha_0 = \beta_0 = 1/2$.

In the case of the ϵ-skew model the alternative priors are

$$\pi_M(\mu, \sigma, \gamma) \propto \frac{1}{\sigma(1-\gamma^2)}, \tag{28}$$

$$\pi_\beta(\mu, \sigma, \gamma) \propto \frac{(1-\gamma)^{\beta_0-1}(1+\gamma)^{\alpha_0-1}}{\sigma}, \tag{29}$$

for $\gamma \in (-1, 1)$. The modified Jeffreys prior does not integrate in γ (like the Jeffreys prior), and only coincides with the AG beta prior in the limit as both α_0 and β_0 tend to zero. This could be argued to be a rather counterintuitive prior on AG, putting lots of mass at the extremes.

The alternative priors for the logistic AG parameterization of Example 1 are

$$\pi_M(\mu, \sigma, \gamma) \propto \frac{1}{\sigma}\,\mathrm{sech}(\gamma)^2, \tag{30}$$
$$\pi_\beta(\mu, \sigma, \gamma) \propto \frac{1}{\sigma}\frac{\left(1 + e^{2\gamma}\right)^{\alpha_0}\left(1 + e^{-2\gamma}\right)^{\beta_0}}{[1 + \cosh(2\gamma)]^{\alpha_0+\beta_0}}, \tag{31}$$

for $\gamma \in \mathbb{R}$. As mentioned above, for $\alpha_0 = \beta_0 = 1$ both priors coincide. Figure 2 shows the graph of the density $\pi_\beta(\gamma)$ corresponding to three parameterizations with $\alpha_0 = \beta_0 = 1$.

Since the modified Jeffreys prior $\pi_M(\cdot)$ is not the Jeffreys prior, the parameterization matters. Whenever the two alternative priors coincide in the examples above, $\pi_M(\cdot)$ corresponds to a symmetric prior in AG, which could be considered "vague" in a rather intuitive sense except for the ϵ-skew case, where the modified Jeffreys prior implies a rather extreme prior when viewed in terms of AG.

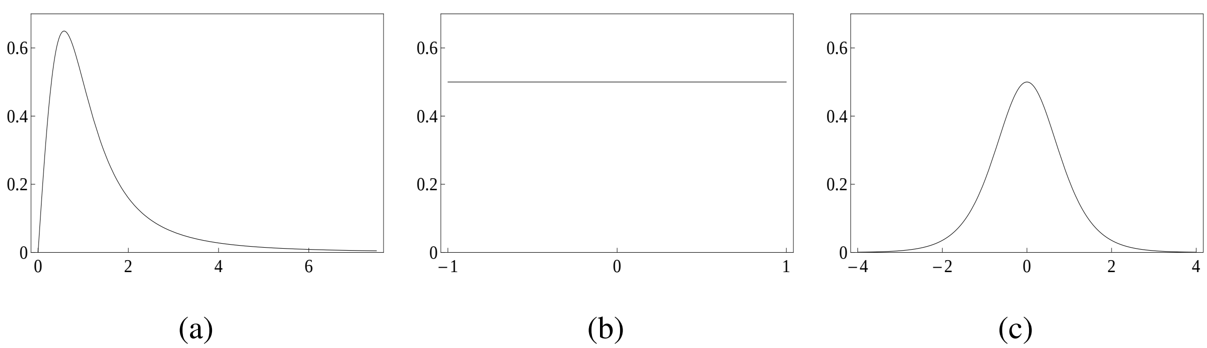


Figure 2: Densities $\pi_\beta(\gamma)$ with $\alpha_0 = \beta_0 = 1$ and: (a) Inverse scale factors parameterization ($\gamma \in \mathbb{R}_+$); (b) $\epsilon-$skew parameterization ($\gamma \in (-1, 1)$); and (c) Logistic AG parameterization ($\gamma \in \mathbb{R}$).

3.4.3 Inference

Since the alternative prior structures are of the form (23), Theorem 6 presents necessary and sufficient conditions for the properness of the posterior distribution of (μ, σ, γ).

Corollary 7 *Consider sampling from (9) where f is a scale mixture of normals. For the Inverse Scale Factors and the logistic AG models the posterior distribution of (μ, σ, γ) using the modified Jeffreys priors (26) and (30), respectively, is proper if $n \geq 2$ and all the observations are different. If $k > 1$ is the largest number of repeated observations in the sample, we have a proper posterior if the mixing distribution of f also satisfies (18).*

Proof. *Follows from Theorem 6(ii) and (iii) given that these priors imply a proper $\pi(\gamma)$.*

The following corollary illustrates that when using the modified Jeffreys prior, the choice of the functions $\{a(\gamma), b(\gamma)\}$ is critical.

Corollary 8 *The posterior distribution under the modified Jeffreys prior (28) in the sampling model (9) with f a scale mixture of normals is improper for the ϵ-skew model.*

Proof. *In this case, the necessary condition (19) is not satisfied.*

However, for the AG beta prior all three model specifications considered here lead to proper posteriors. In fact, posterior existence is guaranteed within a large class of parameterizations $\{a(\gamma), b(\gamma)\}$, namely all parameterizations for which γ is a one-to-one transformation of AG.

Theorem 9 *Let $y = \{y_1, ..., y_n\}$ be a sample from (9) where f is a scale mixture of normals. Consider the AG beta prior in (23) and (25) with $\alpha_0, \beta_0 > 0$. Then, for any choice $\{a(\gamma), b(\gamma)\}$ such that $\lambda(\gamma)$ defined in (8) does not change sign over $\gamma \in \Gamma$ the posterior distribution of (μ, σ, γ) is proper if $n \geq 2$ and all the observations are different. If $k > 1$ is the largest number of repeated observations in the sample, we have a proper posterior if the mixing distribution of f also satisfies (18).*

This result means that for all parameterizations for which γ can be considered a skewness parameter (*i.e.* all choices of $\{a(\gamma), b(\gamma)\}$ of practical modelling interest), we will be able to conduct Bayesian inference with the AG beta prior.

4 Example

Consider the problem of estimating $\theta = P(X < Y)$. The case when X and Y are independent normal or exponential distributions has been recently studied, using Jeffreys priors, by Ventura and Racugno (2011). Now, suppose that X and Y are independent variables from univariate two-piece location-scale models as in (9) with parameters $(\mu_x, \sigma_x, \gamma_x)$ and $(\mu_y, \sigma_y, \gamma_y)$ respectively. We use the data presented in Heinz et al. (2003). This data set contains the body mass index (BMI) of 260 women and 247 men, who are physically active with ages ranging in the twenties and early thirties. Figure 3 shows the histograms of females and males separately. The shape of the histograms suggests the presence of skewness. Therefore, we model these observations with (9), using a normal f.

It has been noted that BMI presents a sexual dimorphism and that men tend to have larger BMI than women. Here, we explore this idea through the posterior distribution of θ. We use the following six models: Model 1 consists of the two-piece model (2) and the independence Jeffreys prior (6). Model 2 corresponds to (9) using $\{a(\gamma), b(\gamma)\}$ of the ϵ-skew model under the independence Jeffreys prior. Model 3 is the logistic AG model of Example 1 for $\gamma \in [-B, B]$ with the Jeffreys prior in (22). Model 4 is the ISF model with the modified Jeffreys prior in (26). Model 5 is the ϵ-skew model in combination with the AG beta prior in (29) and $\alpha_0 = \beta_0 = 1$, which corresponds to a uniform prior on the quantity of interest AG (the posterior distribution of θ is very similar for this model if the hyperparameters are scaled by a factor of 5 or $1/5$). Finally, Model 6 is the skew-normal model of Azzalini (1985), given by

$$s(y|\mu, \sigma, \lambda) = \frac{2}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)\Phi\left(\lambda\frac{y-\mu}{\sigma}\right),$$

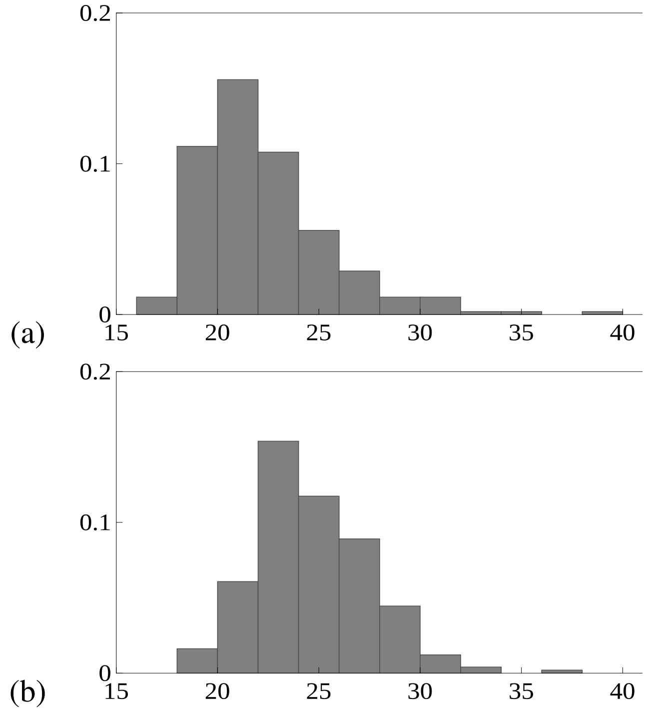


Figure 3: Histograms of body mass index data: (a) females; (b) males.

using the prior

$$\pi(\mu, \sigma, \lambda) \propto \sigma^{-1}\pi(\lambda). \tag{32}$$

The structure of this prior, using the Jeffreys prior of λ derived in the model without location and scale parameters for $\pi(\lambda)$, was proposed in Liseo and Loperfido (2006), who also prove existence of the posterior under this prior. Bayes and Branco (2007) show that the Jeffreys prior of λ can be approximated by a Student t distribution with $1/2$ degrees of freedom, which is what was used for our calculations.

Using a Markov chain Monte Carlo algorithm, a sample of size $10,000$ was recorded from the posterior distribution after a burn-in period of $50,000$ draws with a thinning of 100 draws for all models. Figure 4 presents the posterior distributions of θ.

Clearly, inference with all these different models is very similar, with only the Azzalini model (Model 6) leading to slightly different results. None of the 95% posterior credible intervals include the value $\theta = 0.5$ (in fact the 2.5th percentile is 0.68 for all models), which

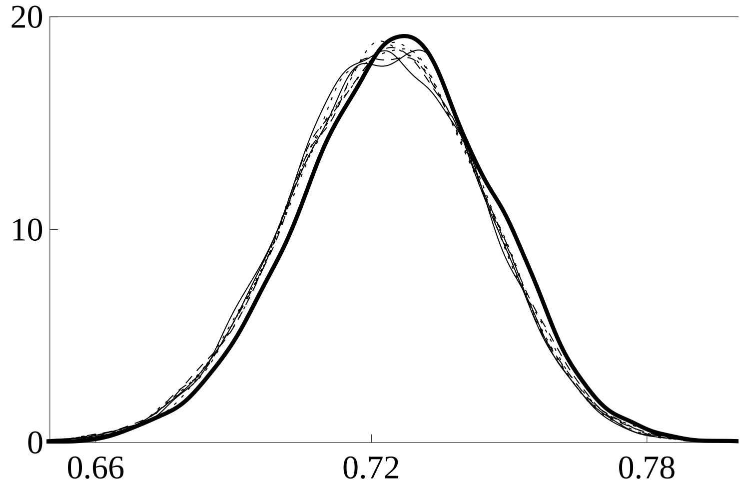


Figure 4: Posterior distributions of θ: Models 1 and 2 (continuous lines); Model 3 with $B = 3$, $B = 10$ and $B = 30$ (dotted lines); Models 4 and 5 (dashed lines); Model 6 (bold line).

is in line with the idea that men tend to have larger BMI than women.

5 Concluding Remarks

We consider the class of univariate continuous two-piece distributions, which are often used as modifications of the symmetric location-scale model to allow for skewness, and its reparameterized versions as presented in Arellano-Valle et al. (2005), where we can identify a location, a scale and a skewness parameter. A number of well-known models (the inverse scale factor or ISF model and the ϵ-skew model) correspond to particular choices of this parameterization. In particular, we focus on Bayesian inference in these models using Jeffreys or the independence Jeffreys prior. We prove that these models do not lead to valid posterior inference under Jeffreys prior for any underlying symmetric distribution in the class of scale mixture of normals. As an ad-hoc fix, we show that modifying Jeffreys prior by truncating the support of the skewness parameter can lead to posterior existence. A more fundamental solution is to use the independence Jeffreys prior instead, which is shown to lead to a valid posterior for some parameterizations of these sampling models. However, this is not the case for the ISF model. Two alternative priors are proposed. A modified Jeffreys prior does lead to a posterior for the ISF model, but not for the ϵ-skew model. A second alternative prior is induced by a Beta prior on the AG skewness measure, and is shown to lead to valid inference in a wide class of parameterizations of these models, including the ISF and ϵ-skew models and arguably all models of practical importance. We apply the models, as well as an alternative skewed distribution due to Azzalini (1985), to some real data. For a number of models that lead to valid inference, we compute empirical coverage probabilities of the posterior credible intervals (see the Supplementary material). This indicates a mostly satisfactory behaviour.

It is important to stress that the three-parameter sampling models examined here are quite simple modifications of the standard location-scale model, and that the Jeffreys prior is a very commonly used prior in the absence of subjective prior information. The fact that the combination of these sampling models with a Jeffreys prior does not lead to a proper posterior is somewhat surprising and definitely relevant for statistical practice, as these models seem attractive options to deal with skewed data, and are used frequently in a wide variety of applied contexts. The better properties of the independence Jeffreys prior are in line with statistical folklore: Jeffreys (1961, p. 182) himself preferred this prior for location-scale problems, and in the univariate normal case the independence Jeffreys is a matching prior (Berger and Sun, 2008). Even with this prior, however, problems of posterior existence can occur, depending on which parameterization we choose. Two alternative priors are examined, and we recommend the AG beta prior for use with two-piece distributions as it ensures posterior inference for any parameterization of practical interest and avoids inducing extreme prior beliefs on the easily interpreted AG skewness measure. Using this prior structure we can induce vague or flat priors on the AG measure of skewness, which is a key function of interest of the model parameters in this context (see Seaman III et al., 2012 for a more general discussion of this principle). The AG beta prior is not an objectively obtained prior (even though it has such an interpretation in special cases), but is easily elicited in practice on the basis of a readily interpretable skewness measure.

Supplementary material, Appendix 1: Proofs

Proof of Theorem 1

The first partial derivatives of $\log[s(y|\mu,\sigma,\gamma)]$ are given by

$$\begin{aligned}
\frac{\partial}{\partial\mu}\log[s(y|\mu,\sigma_1,\sigma_2)] &= -\frac{1}{\sigma_1}\frac{f'\left(\frac{y-\mu}{\sigma_1}\right)}{f\left(\frac{y-\mu}{\sigma_1}\right)}I_{(-\infty,\mu)}(y) - \frac{1}{\sigma_2}\frac{f'\left(\frac{y-\mu}{\sigma_2}\right)}{f\left(\frac{y-\mu}{\sigma_2}\right)}I_{[\mu,\infty)}(y),\\
\frac{\partial}{\partial\sigma_1}\log[s(y|\mu,\sigma_1,\sigma_2)] &= -\frac{1}{\sigma_1+\sigma_2} - \frac{y-\mu}{\sigma_1^2}\frac{f'\left(\frac{y-\mu}{\sigma_1}\right)}{f\left(\frac{y-\mu}{\sigma_1}\right)}I_{(-\infty,\mu)}(y),\\
\frac{\partial}{\partial\sigma_2}\log[s(y|\mu,\sigma_1,\sigma_2)] &= -\frac{1}{\sigma_1+\sigma_2} - \frac{y-\mu}{\sigma_2^2}\frac{f'\left(\frac{y-\mu}{\sigma_2}\right)}{f\left(\frac{y-\mu}{\sigma_2}\right)}I_{[\mu,\infty)}(y).
\end{aligned}$$

Then the entries of the Fisher information matrix of (μ,σ_1,σ_2) are given by

$$\begin{aligned}
I_{11} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\mu}\log[s(y|\mu,\sigma_1,\sigma_2)]\right)^2\right] = \frac{2\alpha_1}{\sigma_1\sigma_2},\\
I_{22} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\sigma_1}\log[s(y|\mu,\sigma_1,\sigma_2)]\right)^2\right] = \frac{\alpha_2}{\sigma_1(\sigma_1+\sigma_2)} + \frac{\sigma_2}{\sigma_1(\sigma_1+\sigma_2)^2},\\
I_{33} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\sigma_2}\log[s(y|\mu,\sigma_1,\sigma_2)]\right)^2\right] = \frac{\alpha_2}{\sigma_2(\sigma_1+\sigma_2)} + \frac{\sigma_1}{\sigma_2(\sigma_1+\sigma_2)^2},\\
I_{12} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\mu}\log[s(y|\mu,\sigma_1,\sigma_2)]\right)\left(\frac{\partial}{\partial\sigma_1}\log[s(y|\mu,\sigma_1,\sigma_2)\right)\right] = -\frac{2\alpha_3}{\sigma_1(\sigma_1+\sigma_2)},\\
I_{13} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\mu}\log[s(y|\mu,\sigma_1,\sigma_2)]\right)\left(\frac{\partial}{\partial\sigma_2}\log[s(y|\mu,\sigma_1,\sigma_2)\right)\right] = \frac{2\alpha_3}{\sigma_2(\sigma_1+\sigma_2)},\\
I_{23} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\sigma_1}\log[s(y|\mu,\sigma_1,\sigma_2)]\right)\left(\frac{\partial}{\partial\sigma_2}\log[s(y|\mu,\sigma_1,\sigma_2)\right)\right] = -\frac{1}{(\sigma_1+\sigma_2)^2}.
\end{aligned}$$

$\square$

Proof of Theorem 2

The determinant of the Fisher information matrix is

$$|I(\mu,\sigma_1,\sigma_2)| = \frac{2\alpha_2\left(\alpha_1+\alpha_1\alpha_2-2\alpha_3^2\right)}{\sigma_1^2\sigma_2^2(\sigma_1+\sigma_2)^2}.$$

We will first prove that $\alpha_2 > 0$. From the definition of α_2 it can only be zero if $1 + tf'(t)/f(t) = 0$ whenever $f(t) > 0$. This means that $f(t) = -tf'(t)$ and this only happens if $f(t) = K/t$ for any positive K. The latter, however, is not a probability density function on $\mathbb{R}$. Thus, α_2 can not be zero.

Next, we will prove that $\alpha_1(1 + \alpha_2) > 2\alpha_3^2$. Applying the Cauchy-Schwarz inequality we have $\alpha_1(1+\alpha_2) \geq 2\alpha_3^2$. We will show that this is a strict inequality. The condition in Theorem 2 implies that

$$0 < \int_0^\infty t \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt.$$

Let

$$\phi(t) = \left|\frac{f'(t)}{\sqrt{f(t)}}\right| > 0 \text{ a.e. and } \psi(t) = t\left|\frac{f'(t)}{\sqrt{f(t)}}\right| > 0 \text{ a.e.}$$

Note that $[\beta\phi(t) + \psi(t)]^2 > 0$ a.e. for any $\beta \in \mathbb{R}$, and thus

$$0 < \int_0^\infty [\beta\phi(t) + \psi(t)]^2\, dt = \beta^2 \int_0^\infty \phi^2(t)\, dt + 2\beta \int_0^\infty \phi(t)\psi(t)\, dt + \int_0^\infty \psi^2(t)\, dt.$$

This is a polynomial of degree 2 in β with positive coefficients and no real roots, implying that the discriminant is negative, so that

$$\left[\int_0^\infty t \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt\right]^2 < \left[\int_0^\infty t^2 \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt\right] \left[\int_0^\infty \left[\frac{f'(t)}{f(t)}\right]^2 f(t)\, dt\right].$$

$\square$

Proof of Theorem 3

The first partial derivatives of $\log[s(y|\mu,\sigma,\gamma)]$ are given by

$$\begin{aligned}
\frac{\partial}{\partial\mu}\log[s(y|\mu,\sigma,\gamma)] &= -\frac{1}{\sigma b(\gamma)}\frac{f'\left(\frac{y-\mu}{\sigma b(\gamma)}\right)}{f\left(\frac{y-\mu}{\sigma b(\gamma)}\right)}I_{(-\infty,\mu)}(y) - \frac{1}{\sigma a(\gamma)}\frac{f'\left(\frac{y-\mu}{\sigma a(\gamma)}\right)}{f\left(\frac{y-\mu}{\sigma a(\gamma)}\right)}I_{[\mu,\infty)}(y),\\
\frac{\partial}{\partial\sigma}\log[s(y|\mu,\sigma,\gamma)] &= -\frac{1}{\sigma} - \frac{y-\mu}{\sigma^2 b(\gamma)}\frac{f'\left(\frac{y-\mu}{\sigma b(\gamma)}\right)}{f\left(\frac{y-\mu}{\sigma b(\gamma)}\right)}I_{(-\infty,\mu)}(y) - \frac{y-\mu}{\sigma^2 a(\gamma)}\frac{f'\left(\frac{y-\mu}{\sigma a(\gamma)}\right)}{f\left(\frac{y-\mu}{\sigma a(\gamma)}\right)}I_{[\mu,\infty)}(y),\\
\frac{\partial}{\partial\gamma}\log[s(y|\mu,\sigma,\gamma)] &= -\frac{a'(\gamma)+b'(\gamma)}{a(\gamma)+b(\gamma)} - \frac{y-\mu}{\sigma}\frac{b'(\gamma)}{b(\gamma)^2}\frac{f'\left(\frac{y-\mu}{\sigma b(\gamma)}\right)}{f\left(\frac{y-\mu}{\sigma b(\gamma)}\right)}I_{(-\infty,\mu)}(y)\\
&- \frac{y-\mu}{\sigma}\frac{a'(\gamma)}{a(\gamma)^2}\frac{f'\left(\frac{y-\mu}{\sigma a(\gamma)}\right)}{f\left(\frac{y-\mu}{\sigma a(\gamma)}\right)}I_{[\mu,\infty)}(y).
\end{aligned}$$

Thus, the entries of the Fisher information matrix of (μ,σ,γ) are

$$\begin{aligned}
I_{11} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\mu}\log[s(y|\mu,\sigma,\gamma)]\right)^2\right] = \frac{2\alpha_1}{a(\gamma)b(\gamma)\sigma^2},\\
I_{22} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\sigma}\log[s(y|\mu,\sigma,\gamma)]\right)^2\right] = \frac{\alpha_2}{\sigma^2},\\
I_{33} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\gamma}\log[s(y|\mu,\sigma,\gamma)]\right)^2\right] = \frac{\alpha_2+1}{a(\gamma)+b(\gamma)}\left[\frac{b'(\gamma)^2}{b(\gamma)}+\frac{a'(\gamma)^2}{a(\gamma)}\right] - \left[\frac{a'(\gamma)+b'(\gamma)}{a(\gamma)+b(\gamma)}\right]^2,\\
I_{12} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\mu}\log[s(y|\mu,\sigma,\gamma)]\right)\left(\frac{\partial}{\partial\sigma}\log[s(y|\mu,\sigma,\gamma)\right)\right] = 0,\\
I_{13} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\mu}\log[s(y|\mu,\sigma,\gamma)]\right)\left(\frac{\partial}{\partial\gamma}\log[s(y|\mu,\sigma,\gamma)\right)\right]\\
&= \frac{2\alpha_3}{\sigma[a(\gamma)+b(\gamma)]}\left[\frac{a'(\gamma)}{a(\gamma)} - \frac{b'(\gamma)}{b(\gamma)}\right],\\
I_{23} &= \mathbb{E}\left[\left(\frac{\partial}{\partial\sigma}\log[s(y|\mu,\sigma,\gamma)]\right)\left(\frac{\partial}{\partial\gamma}\log[s(y|\mu,\sigma,\gamma)\right)\right]\\
&= \frac{\alpha_2}{\sigma}\left[\frac{a'(\gamma)+b'(\gamma)}{a(\gamma)+b(\gamma)}\right].
\end{aligned}$$

□

Proof of Theorem 4

Note that

$$\frac{d}{d\gamma}AG(\gamma) = 2\frac{a'(\gamma)b(\gamma) - a(\gamma)b'(\gamma)}{[a(\gamma)+b(\gamma)]^2} = 2\frac{a(\gamma)b(\gamma)\lambda(\gamma)}{[a(\gamma)+b(\gamma)]^2},$$

so that

$$\frac{dAG(\gamma)}{d\gamma} > 0 \Leftrightarrow \lambda(\gamma) > 0 \text{ and } \frac{dAG(\gamma)}{d\gamma} < 0 \Leftrightarrow \lambda(\gamma) < 0.$$

$\square$

Proof of Theorem 5

First of all, consider the independence Jeffreys prior (6) and the change of variable (7), then

$$\begin{aligned}\pi_I(\mu,\sigma,\gamma) &\propto \frac{|a'(\gamma)b(\gamma) - a(\gamma)b'(\gamma)|\sqrt{[b(\gamma)+\alpha_2[a(\gamma)+b(\gamma)]][a(\gamma)+\alpha_2[a(\gamma)+b(\gamma)]]}}{\sigma\sqrt{a(\gamma)b(\gamma)}[a(\gamma)+b(\gamma)]^2} \\ &\leq \frac{(\alpha_2+1)|a'(\gamma)b(\gamma) - a(\gamma)b'(\gamma)|}{\sigma\sqrt{a(\gamma)b(\gamma)}[a(\gamma)+b(\gamma)]}.\end{aligned}$$

For the particular choice $\{a(\gamma), b(\gamma)\} = \{\gamma, 1/\gamma\}$, the upper bound of $\pi_I(\mu,\sigma,\gamma)$ is proportional to $[\sigma(1+\gamma^2)]^{-1}$. Now, the proof of (i) and (ii) is as follows.

(i) Applying Theorem 1 from Fernández and Steel (1999) and using this upper bound we can derive the properness of the posterior distribution of (μ,σ,γ). Now, since the mapping $(\mu,\sigma,\gamma) \leftrightarrow (\mu,\sigma_1,\sigma_2)$ is one-to-one, it follows that the posterior distribution of (μ,σ_1,σ_2) is proper.

(ii) The proof follows analogously by applying Theorem 2 from Fernández and Steel (1999).

$\square$

Proof of Theorem 6

Let f be a scale mixture of normals with τ_j the mixing variable associated with y_j and where the τ_js are independent random variables defined on $\mathbb{R}^+$ with distribution P_{τ_j}.

(i) Integrating with respect of μ over a subspace we get a lower bound for the marginal distribution of $(y_1, ..., y_n)$ which is proportional to

$$\int_{\mathbb{R}_+^n}\int_\Gamma\int_0^\infty\int_{-\infty}^{y_{(1)}}\left(\prod_{j=1}^n \tau_j^{\frac{1}{2}}\right)\frac{\sigma^{-(n+1)}}{[a(\gamma)+b(\gamma)]^n}\exp\left[-\frac{1}{2\sigma^2 a(\gamma)^2}\sum_{j=1}^n \tau_j(y_j-\mu)^2\right]$$
$$\times \quad \pi(\gamma)\,d\mu d\sigma d\gamma dP_{(\tau_1,...,\tau_n)}.$$

Using the change of variable $\vartheta = \sigma a(\gamma)$, we can rewrite the lower bound as follows

$$\int_\Gamma\left[\frac{a(\gamma)}{a(\gamma)+b(\gamma)}\right]^n\pi(\gamma)\,d\gamma\int_{\mathbb{R}_+^n}\int_0^\infty\int_{-\infty}^{y_{(1)}}\left(\prod_{j=1}^n \tau_j^{\frac{1}{2}}\right)\vartheta^{-(n+1)}$$
$$\times \quad \exp\left[-\frac{1}{2\vartheta^2}\sum_{j=1}^n \tau_j(y_j-\mu)^2\right]\,d\mu d\vartheta dP_{(\tau_1,...,\tau_n)},$$

and the result follows.

(ii) We can get an upper bound for the marginal distribution of $(y_1, ..., y_n)$ proportional to

$$\int_{\mathbb{R}_n^+}\int_\Gamma\int_0^\infty\int_{-\infty}^{\infty}\left(\prod_{j=1}^n \tau_j^{\frac{1}{2}}\right)\frac{\sigma^{-(n+1)}}{[a(\gamma)+b(\gamma)]^n}\exp\left[-\frac{1}{2\sigma^2 h(\gamma)^2}\sum_{j=1}^n \tau_j(y_j-\mu)^2\right]$$
$$\times \quad \pi(\gamma)\,d\mu d\sigma d\gamma dP_{(\tau_1,...,\tau_n)},$$

where $h(\gamma) = \max\{a(\gamma), b(\gamma)\}$. Consider the change of variable $\vartheta = \sigma h(\gamma)$ and rewrite the upper bound as follows

$$\int_\Gamma\left[\frac{h(\gamma)}{a(\gamma)+b(\gamma)}\right]^n\pi(\gamma)\,d\gamma\int_{\mathbb{R}_n^+}\int_0^\infty\int_{-\infty}^{\infty}\left(\prod_{j=1}^n \tau_j^{\frac{1}{2}}\right)\vartheta^{-(n+1)}$$
$$\times \quad \exp\left[-\frac{1}{2\vartheta^2}\sum_{j=1}^n \tau_j(y_j-\mu)^2\right]\,d\mu d\vartheta dP_{(\tau_1,...,\tau_n)}.$$

Fernández and Steel (2000, Th. 1) show that the integral in $\mu, \vartheta, \tau_1, ..., \tau_n$ is finite if $n \geq 2$. Then, by Theorem 1 from Fernández and Steel (1999), the existence of the integral in γ is a sufficient condition for the properness of the posterior distribution of (μ, σ, γ). The result then follows from

$$\int_\Gamma\left[\frac{h(\gamma)}{a(\gamma)+b(\gamma)}\right]^n\pi(\gamma)\,d\gamma \leq \int_\Gamma \pi(\gamma)\,d\gamma.$$

(iii) The proof follows analogously by applying Theorem 2 from Fernández and Steel (1999).

□

Proof of Theorem 7

If f is a scale mixture of normals, then integrating over a subspace with respect to μ we get a lower bound for the marginal distribution of $(y_1, ..., y_n)$ which is proportional to

$$\int_{\mathbb{R}_+^n} \int_\Gamma \int_0^\infty \int_{-\infty}^{y_{(1)}} \left(\prod_{j=1}^n \tau_j^{\frac{1}{2}}\right) \frac{\sigma^{-(n+2)}}{[a(\gamma)+b(\gamma)]^n} \exp\left[-\frac{1}{2\sigma^2 a(\gamma)^2} \sum_{j=1}^n \tau_j (y_j-\mu)^2\right]$$
$$\times \frac{|\lambda(\gamma)|}{a(\gamma)+b(\gamma)}\, d\mu d\sigma d\gamma dP_{(\tau_1,...,\tau_n)}.$$

Using the change of variable $\vartheta = \sigma a(\gamma)$, we can rewrite this lower bound as follows

$$\int_\Gamma \left[\frac{a(\gamma)}{a(\gamma)+b(\gamma)}\right]^{n+1} |\lambda(\gamma)|\, d\gamma \int_{\mathbb{R}_+^n} \int_0^\infty \int_{-\infty}^{y_{(1)}} \left(\prod_{j=1}^n \tau_j^{\frac{1}{2}}\right) \vartheta^{-(n+2)}$$
$$\times \exp\left[-\frac{1}{2\vartheta^2} \sum_{j=1}^n \tau_j (y_j-\mu)^2\right] d\mu d\vartheta dP_{(\tau_1,...,\tau_n)}.$$

Therefore, the existence of the first integral is a necessary condition for the properness of the posterior distribution of (μ, σ, γ). □

Proof of Theorem 8

The proof of (i) is as follows. If f is normal, defining $h(\gamma) = \max\{a(\gamma), b(\gamma)\}$ we get an upper bound for the marginal distribution of $(y_1, ..., y_n)$ which is proportional to

$$\int_{-\infty}^\infty \int_\Gamma \int_0^\infty \frac{\pi_J(\mu,\sigma,\gamma)}{[a(\gamma)+b(\gamma)]^n \sigma^n} \exp\left[-\frac{1}{2\sigma^2 h(\gamma)^2} \sum_{j=1}^n (y_j-\mu)^2\right] d\sigma d\gamma d\mu$$
$$\propto \int_{-\infty}^\infty \left[\sum_{j=1}^n (y_j-\mu)^2\right]^{-\frac{n+1}{2}} d\mu \int_\Gamma \frac{h(\gamma)^{n+1}}{[a(\gamma)+b(\gamma)]^{n+1}} |\lambda(\gamma)| d\gamma.$$

The first integral exists if $n \geq 2$ and at least 2 observations are different. Then the existence of the second integral is a sufficient condition for the existence of the posterior distribution. For the second integral we use that

$$\int_{\Gamma} \frac{h(\gamma)^{n+1}}{[a(\gamma)+b(\gamma)]^{n+1}} |\lambda(\gamma)| d\gamma \leq \int_{\Gamma} |\lambda(\gamma)| d\gamma,$$

which is finite by assumption. If f is Laplace, analogously to the normal case we get an upper bound for the marginal distribution of $(y_1, ..., y_n)$ which is proportional to

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{\Gamma} \int_{0}^{\infty} \frac{\pi_J(\mu, \sigma, \gamma)}{[a(\gamma)+b(\gamma)]^n \sigma^n} \exp\left[-\frac{1}{\sigma h(\gamma)} \sum_{j=1}^{n} |y_j - \mu| \right] d\sigma d\gamma d\mu \\ \propto \quad & \int_{-\infty}^{\infty} \left[\sum_{j=1}^{n} |y_j - \mu| \right]^{-(n+1)} d\mu \int_{\Gamma} \frac{h(\gamma)^{n+1}}{[a(\gamma)+b(\gamma)]^{n+1}} |\lambda(\gamma)| d\gamma, \end{aligned}$$

and the same argument leads to the result.

Result (ii) follows immediately from Corollary 6.

For (iii) let us assume, without loss of generality, that $AG(\gamma)$ is an increasing function and $\Gamma = (\underline{\gamma}, \overline{\gamma})$. First, note that we can rewrite $AG(\gamma)$ as follows

$$AG(\gamma) = \tanh\left\{ \frac{1}{2} \log\left[\frac{a(\gamma)}{b(\gamma)} \right] \right\}.$$

Then

$$\begin{aligned} \lim_{\gamma \to \overline{\gamma}} AG(\gamma) = 1 \quad & \Leftrightarrow \quad \lim_{\gamma \to \overline{\gamma}} \log\left[\frac{a(\gamma)}{b(\gamma)} \right] = \infty \\ \lim_{\gamma \to \underline{\gamma}} AG(\gamma) = -1 \quad & \Leftrightarrow \quad \lim_{\gamma \to \underline{\gamma}} \log\left[\frac{a(\gamma)}{b(\gamma)} \right] = -\infty, \end{aligned}$$

which contradicts the assumption that $\lambda(\gamma)$ is absolutely integrable. The result is analogous if AG is decreasing. $\square$

Proof of Theorem 9

From Theorem 6(ii) and (iii) we know that properness of $\pi(\gamma)$ in (23) is sufficient for existence of the posterior. The AG beta prior implies a proper prior for AG when $\alpha_0, \beta_0 > 0$. From

Theorem 4 the condition that $\lambda(\gamma)$ does not change sign is equivalent to AG being a one-to-one transformation of γ. Thus, the induced prior on γ will be proper and the result follows. $\square$

Supplementary material, Appendix 2: Simulation Study

In this section we investigate the empirical coverage of the 95% posterior credible intervals, defined by the 2.5th and 97.5th percentiles. We simulate $N = 10,000$ data sets of size $n = 30, 100$ and 1000 from seven sampling models, Models 1-5 described in Section 4 plus two additional models described below, where we take f to be a normal distribution throughout, and analyse these data using the corresponding Bayesian model. Model 7 corresponds to the Logistic AG model model with AG beta prior and $\alpha_0 = \beta_0 = 1$, and Model 8 consists of the Inverse scale factors model with AG beta prior and $\alpha_0 = \beta_0 = 1$. For each of these N datasets, a sample of size $3,000$ was obtained from the posterior distribution using a Markov chain Monte Carlo sampler after a burn-in period of $5,000$ iterations and thinned to every 50th iteration. Finally, the proportion of 95% credible intervals that include the true value of the parameter was calculated. Results are presented in Tables 1-7. For Model 3 we know that the truncation to a finite interval is what makes the posterior well-defined. To investigate how sensitive the results are to the particular value chosen for B, we have experimented with various values. Models 5, 7 and 8 employ the same sort of prior with different parameterizations of the sampling model (9), while Models 1–4 differ in both the kind of prior employed and the parameterization of the sampling model.

All models lead to coverage probabilities above the nominal level for samples of size $n = 30$, especially in the case of σ for Models 3–5 and 7. Once we increase the sample size to $n = 100$, the coverage is quite close to the nominal value, except for one setting for Model 1, where the coverage is still a bit high. As we further increase to samples of 1000 observations, all cases lead to coverage very close to 95%, as we would expect. The simulation standard

Sample size	$n = 30$		$n = 100$		$n = 1000$	
Parameters	$\sigma_1 = 2.0$ $\sigma_2 = 0.5$	$\sigma_1 = 0.66$ $\sigma_2 = 1.50$	$\sigma_1 = 2.0$ $\sigma_2 = 0.5$	$\sigma_1 = 0.66$ $\sigma_2 = 1.50$	$\sigma_1 = 2.0$ $\sigma_2 = 0.5$	$\sigma_1 = 0.66$ $\sigma_2 = 1.50$
μ	0.976	0.967	0.971	0.956	0.948	0.953
σ_1	0.961	0.951	0.974	0.958	0.947	0.949
σ_2	0.975	0.971	0.961	0.951	0.948	0.950

Table 1: Coverage proportions. Two-piece model in (2) with independence Jeffreys prior (Model 1)

Sample size	$n = 30$		$n = 100$		$n = 1000$	
Parameter	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$
μ	0.971	0.967	0.954	0.955	0.947	0.948
σ	0.959	0.960	0.947	0.945	0.953	0.954
γ	0.971	0.969	0.957	0.957	0.948	0.952

Table 2: Coverage proportions. ϵ-skew model with independence Jeffreys prior (Model 2)

Sample size	$n = 30$		$n = 100$		$n = 1000$	
Parameter	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$
μ	0.967	0.964	0.949	0.953	0.948	0.949
σ	0.995	0.991	0.952	0.960	0.948	0.947
γ	0.964	0.965	0.949	0.952	0.948	0.947

Table 3: Coverage proportions. Logistic AG model with Jeffreys prior (Model 3) and $B = 3$

errors are around 0.002 for all cases, so that for large n most differences in the tables can simply be accounted for by Monte Carlo error. For Model 3, the choice of B (we have also tried $B = 10$ and $B = 30$) did not seem to have any noticeable effect. Comparing Tables 2 and 5, Tables 3 and 6 and Tables 4 and 7 allows us to assess the difference in coverage between the AG beta prior and the other priors, and we can conclude these differences are quite small. The

Size	$n = 30$		$n = 100$		$n = 1000$	
Parameter	$\gamma = 0.5$	$\gamma = 1.5$	$\gamma = 0.5$	$\gamma = 1.5$	$\gamma = 0.5$	$\gamma = 1.5$
μ	0.969	0.967	0.963	0.950	0.949	0.946
σ	0.992	0.972	0.965	0.949	0.947	0.949
γ	0.967	0.971	0.967	0.950	0.950	0.948

Table 4: Coverage proportions: Inverse scale factors model with modified Jeffreys prior (Model 4)

Size	$n = 30$		$n = 100$		$n = 1000$	
Parameter	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$
μ	0.968	0.967	0.960	0.959	0.947	0.951
σ	0.994	0.993	0.968	0.970	0.947	0.951
γ	0.968	0.969	0.964	0.964	0.948	0.950

Table 5: Coverage proportions: ϵ-skew model with AG beta prior and $\alpha_0 = \beta_0 = 1$ (Model 5).

Size	$n = 30$		$n = 100$		$n = 1000$	
Parameter	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$
μ	0.965	0.965	0.956	0.961	0.945	0.950
σ	0.992	0.994	0.964	0.966	0.950	0.952
γ	0.968	0.968	0.960	0.965	0.947	0.948

Table 6: Coverage proportions: Logistic AG model with AG beta prior and $\alpha_0 = \beta_0 = 1$ (Model 7).

only exception is the performance for σ with 30 observations from the ϵ-skew model, where the independence Jeffreys prior leads to better coverage. Overall, the frequentist coverage properties of the models examined are pretty good, with perhaps Model 2 displaying the best performance.

We also conducted the same simulation study using a skewed version of a Student-t sampling model with 2 degrees of freedom and we observed a rather similar behaviour of the

Size	$n = 30$		$n = 100$		$n = 1000$	
Parameter	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$	$\gamma = 0.5$	$\gamma = -0.5$
μ	0.969	0.973	0.952	0.949	0.949	0.948
σ	0.986	0.973	0.963	0.953	0.950	0.951
γ	0.968	0.976	0.959	0.951	0.946	0.951

Table 7: Coverage proportions: Inverse scale factors model with AG beta prior and $\alpha_0 = \beta_0 = 1$ (Model 8).

coverage proportions. Interestingly, however, the coverage for the ϵ-skew model with $n = 30$ is better in this case with the AG beta prior than under the independence Jeffreys and the overall coverage for σ in small samples is better than with the skewed normal throughout.

References

Arnold, B. C. and Groeneveld, R. A. (1995). Measuring skewness with respect to the mode. *The American Statistician* 49: 34–38.

Arellano-Valle, R. B., Gómez, H. W. and Quintana, F. A. (2005). Statistical inference for a general class of asymmetric distributions. *Journal of Statistical Planning and Inference* 128: 427–443.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12: 171-178.

Bayes, C. L. and Branco, M. D. (2007). Bayesian inference for the skewness parameter of the scalar skew-normal distribution. *Brazilian Journal of Probability and Statistics* 21: 141–163.

Berger, J. O., Bernardo, J. M. and Sun, D. (2009). The formal definition of reference priors. *Annals of Statistics* 37: 905–938.

Berger, J. O. and Sun, D. (2008). Objective priors for the bivariate normal model. *Annals of Statistics* 36: 963–982.

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Reading, Mass: Addison-Wesley.

Clarke, B. and Barron, A. R. (1994). Jeffreys' prior is asymptotically least favorable under entropy risk. *Journal of Statistical Planning and Inference* 41: 37–60.

Cox, D. R. and Reid, N. (1987). Orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society*, Series B 49: 1–39.

Fernández, C. and Steel, M. F. J. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* 93: 359–371.

Fernández, C. and Steel, M. F. J. (1999). On the dangers of modelling through continuous distributions: a Bayesian perspective. In *Bayesian Statistics* 6. J. O. Berger, J. M. Bernardo, A.P. Dawid and A. F. M. Smith, eds. Oxford University Press, Oxford, pp. 213-238 (with discussion).

Fernández, C. and Steel, M. F. J. (2000). Bayesian regression analysis with scale mixtures of normals. *Econometric Theory* 16: 80–101.

Fonseca, T. C. O., Ferreira, M. A. R and Migon, H. S. (2008). Objective Bayesian analysis for the Student-t regression model. *Biometrika* 95: 325–333.

Galbraith, J.W. and van Norden, S. (2012). Assessing gross domestic product and inflation forecasts derived from Bank of England fan charts. *Journal of the Royal Statistical Society, A* 175: 713–727.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 57: 1317-1339.

Gibbons, J. F. and Mylroie, S. (1973). Estimation of impurity profiles in ion-implanted amorphous targets using joined half-Gaussian distributions. *Applied Physics Letters* 22: 568–569.

Heinz G., Peterson, L. J., Johnson, R. W. and Kerk, C.J. (2003). Exploring relationships in body dimensions. *Journal of Statistics Education* (online only) 11(2). URL: www.amstat.org/publications/jse/v11n2/datasets.heinz.html.

Ibrahim, J. G. and Laud, P. W. (1991). On Bayesian analyisis of generalized linear models using Jeffreys's priors. *Journal of the American Statistical Association* 86: 981–986.

Jeffreys, H. (1941). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London.* Series A, 183: 453–461.

Jeffreys, H. (1961). *Theory of Probability* (3rd ed.) Oxford: Clarendon.

John, S. (1982). The three-parameter two-piece normal family of distributions and its fitting. *Communications in Statistics - Theory and Methods* 11: 879-885.

Jones, M. C. (2006). A note on rescalings, reparametrizations and classes of distributions. *Journal of Statistical Planning and Inference* 136: 3730-3733.

Jones, M. C. and Anaya-Izquierdo K. (2011). On parameter orthogonality in symmetric and skew models. *Journal of Statistical Planning and Inference* 141: 758–770.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation.* Springer, New York.

Liseo, B. and Loperfido, N. (2006). A note on reference priors for the scalar skew-normal distribution. *Journal of Statistical Planning and Inference* 136: 373-389.

Mudholkar, G. S. and Hutson, A. D. (2000). The epsilon-skew-normal distribution for analyzing near-normal data. *Journal of Statistical Planning and Inference* 83: 291–309.

Punathumparambath B., Kulathinal, S. and George, S. (2012). Asymmetric type II compound Laplace distribution and its application to microarray gene expression. *Computational Statistics and Data Analysis* 56: 1396–1404.

Purdom, E. and Holmes, S.P (2005). Error distribution for gene expression data. *Statistical Applications in Genetics and Molecular Biology* 4: 16.

Rubio, F. J. and Steel, M. F. J. (2011). Inference for grouped data with a truncated skew-Laplace distribution. *Computational Statistics and Data Analysis* 55: 3218-3231.

Seaman III, J. W., Seaman Jr., J. W. and Stamey, J. D. (2012). Hidden dangers of specifying noninformative priors. *The American Statistician* 66: 77–84.

Trindade, A. A., Zhu, Y. and Andrews, B. (2010). Time series models with asymmetric Laplace innovations. *Journal of Statistical Computation and Simulation* 80: 1317–1333.

Ventura, L. and Racugno, W. (2011). Recent advances on Bayesian inference for $P(X < Y)$. *Bayesian Analysis* 6: 1–18.

Wallis, K. (2004). An assessment of Bank of England and National Institute inflation forecast uncertainties. *National Institute Economic Review* 189: 64–71.

Wallis, K. (2013). The two-piece normal, binormal, or double Gaussian distribution: its origin and rediscoveries. *Statistical Science*, forthcoming.